

Data Science and the Art of Producing Entertainment at Netflix



Netflix Technology Blog

[Follow](#)

Mar 26, 2018 · 10 min read

by [Ritwik Kumar](#), [Vinit Misra](#), [Jen Walraven](#), [Lavanya Sharan](#),
[Bahareh Azarnoush](#), [Boris Chen](#), [Nirmal Govind](#)

Netflix has released hundreds of Originals and plans to spend \$8 billion over the next year on content. Creators of these stories pour their hearts and souls into turning ideas into joy for our viewers. The sublime art of doing this well is hard to describe, but it necessitates a careful orchestration of creative, business and technical decisions. Here we will focus on the latter two—business & technical decisions like planning budgets, finding locations, building sets, and scheduling guest actors that enable the creative act of connecting with viewers.



Each production is a mountain of operational and logistical challenges that consumes and produces tremendous amounts of data. At Netflix's scale, this is further amplified to levels seldom encountered before in the history of entertainment. This has created opportunities to organize, analyze and model this data that are equally singular in history. This is where data science can aid the art of producing entertainment.



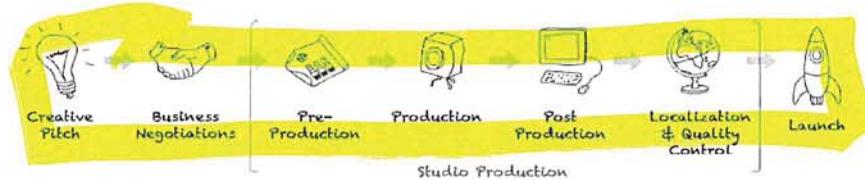


Figure 1. Content Creation Lifecycle

From the moment a show is pitched and before it shows up on our service, it goes through a few broad stages that are depicted in Figure 1. Studio Production refers to Pre-Production (planning, budgeting, etc.), Production (principal photography), Post Production (editing, sound mixing, etc.), Localization & Quality Control (subtitle creation, snuffing out technical glitches, etc.). In the rest of this blog, we will follow a title's journey through these stages and examine some questions that data science can help answer.

Georgia or Gibraltar?

During Pre-Production, producers and executives are tasked with critical decisions such as: do we shoot in Georgia or in Gibraltar? Do we hire a thousand extras or lean on VFX? Do we keep a 10-hour workday or a 12-hour workday? Each of these choices can have massive impact on cost, timeline and creative outcome of the project. Traditionally, these decisions are rooted in human experience and intuition. Let us see how one can supplement these with data derived insights.



Altered Carbon (Photo Credit: Katie Yu / Netflix)

The problem we are describing is one of cost estimation: given various attributes about a production, estimate how much it will cost. These attributes should characterize both the content (genre, similarity to other titles, etc.) as well as the aforementioned production decisions (geographical location, production appetite, schedule, etc.). A production team could use this model as a sandbox to answer the central question of prep: which combination of production decisions stays most true to the creative vision while also staying under budget?

The core challenge with building any such model is data sparsity. A production executive may want to evaluate Atlanta, Georgia as a shooting location for a big budget fantasy epic, but we may not have much historical data about Atlanta. Furthermore, the mechanisms by which location impacts cost may be complex and difficult to infer from data, even in popular shooting locations.

One solution is to lean heavily on domain knowledge and expertise. Rather than attempting to learn thousands of parameters in a black box, we carefully construct networks that reflect our intuition about the problem space, and place strong priors on any parameters we seek to learn.

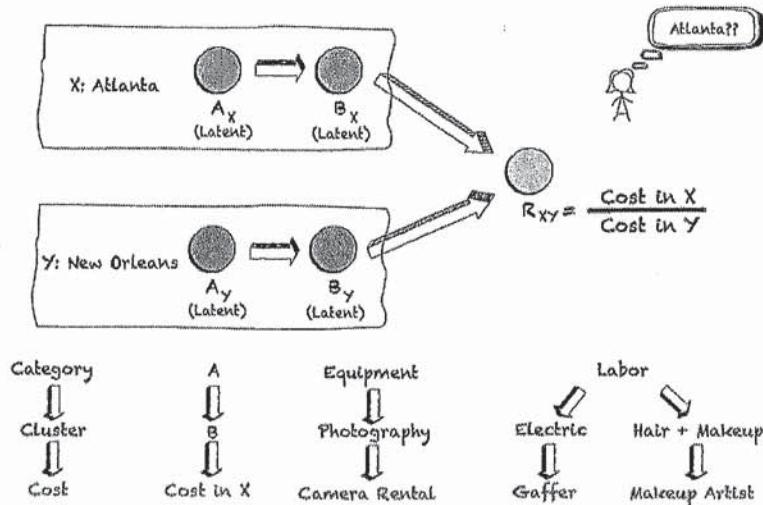


Figure 2. An example hierarchical cost model

For instance, suppose our goal is to model the ratio R_{XY} in certain production costs between geographic locations X (Atlanta) and Y (New Orleans). There are hundreds of such costs for any given production, and some vary more with location than others. Despite the sparsity in

the data, one can efficiently model the fractional change in each of these costs by organizing them hierarchically (illustrated in Figure 2), placing them in a model that reflects this hierarchy, and, finally, putting priors on these ratios that reflect domain expertise. Such a model allows production executives to play around with ‘what-if’ scenarios and make informed decisions about critical aspects of a production.

Once the pre-production minefield of decisions has been navigated, next we get to tackle the challenge of putting a plan into action. This marks the start of the Production stage (Figure 1). This is where it takes sorting through a mountain of logistical and operational challenges to enable the creative act of principal photography. These challenges are perhaps best illustrated by the task of Scheduling, which dictates how various resources are orchestrated for principal photography. We examine Scheduling next.

It's a wrap for the day, or is it?

A multi-episode show can easily have hundreds of scenes, shot with hundreds of on- and off-screen individuals, over dozens of locations, spanning many months. Efficiently and effectively scheduling this symphony—the task of the first assistant director (1st AD)—is the central logistical challenge in film and TV production, and it is as much an art form as the performances one sees on-screen. When done manually, it is not uncommon for a 1st AD to spend hundreds of hours building a schedule.



Marvel's Jessica Jones (Photo Credit: David Giesbrecht / Netflix)

At its core, a schedule is an ordering of scenes for each day of principal photography. A 1st AD's job is to create such an ordering that respects various constraints and objectives: e.g. "We only have Actor X for one week." Many of these considerations are fundamentally human judgments, but there is also room for automation to provide suggestions, or to assist with the more mechanical side of things. In particular, mathematical optimization can help generate rough schedules to inform early-stage production planning.

Let's describe a simplistic model that attempts to roughly capture the most basic scheduling considerations. Suppose our production consists of N scenes being shot by a single unit, over the course of D days, in L locations, and suppose we are given a rough estimate for the time each scene will take to shoot. To formulate this as an optimization problem, we need to specify variables, constraints, and an objective.

Variables:

- The day and time each scene starts shooting.
- For each location and cast member, the day when the contract starts and the day it ends.

Sample Constraints:

- We can only be shooting one scene at any time.
- All resources (location, cast) needed for a scene must be on contract when that scene is shooting.
- A shooting day can't run too long, and the cast and crew needs a minimum amount of time to rest between workdays.
- It takes time to move from one location to another, depending on the locations.
- A strictly daytime scene must be shot before sunset, and a strictly nighttime scene after sunset.

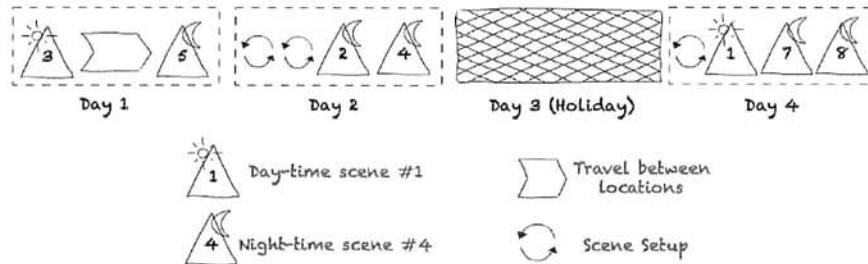


Figure 3: A sample schedule

Objective (very crude approximations!):

- Crew is paid a fixed amount per shoot day.
- Performers and locations are paid a fixed amount (varies between performers and locations) per contractual day.

Even with a simplistic model like this, one can generate reasonable-looking schedules within minutes. Such schedules are also useful for early stage planning (e.g. budgeting) and as starting points for more refined scheduling. For expert users, models of this sort can even be used as interactive tools to selectively adjust portions of a schedule.

After principal photography is complete, for a typical show, there could still be 100 to 200 tasks like editing, sound mixing, color correction etc. that need to be completed before it is ready for the screen.

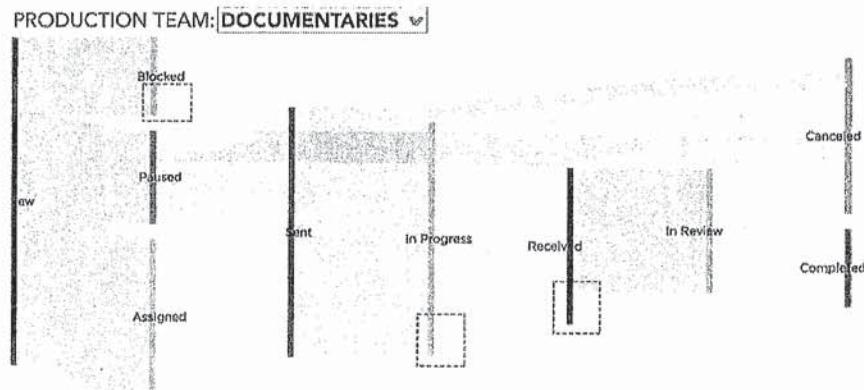
Coordination and tracking of these tasks is tackled in Post Production (Figure 1). For many TV shows and films, Post Production easily ends up consuming far more time than principal photography. For instance,

'Apocalypse Now' famously took over two years of post production before it was ready for the screens. Let us examine how data science can help Post Production next.

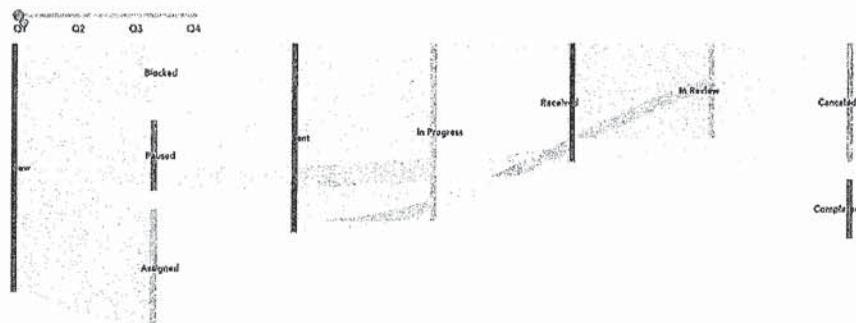
Visuals Behind the Scenes

As Post Production teams track hundreds of tasks per TV show or film, identifying bottlenecks and blockers that span multiple titles becomes an even greater challenge. Leveraging data science to equip teams with the ability to slice, dice and visualize this data at scale can help identify anomalies and opportunities for interventions in their complex, multilayered cross-team workflows.

As an illustration, let us look at the Sankey diagram below representing a generic post production asset workflow. This captures the progression of assets—like VFX shots, daily film clips, or final cuts—through a review process. In this example things flow pretty nicely, but we see drop-offs(red boxes) at the Blocked, In Progress, and Received stages. With this detail, teams can dive in to understand what is blocked or undelivered, using additional tools to evaluate these gaps on specific productions.

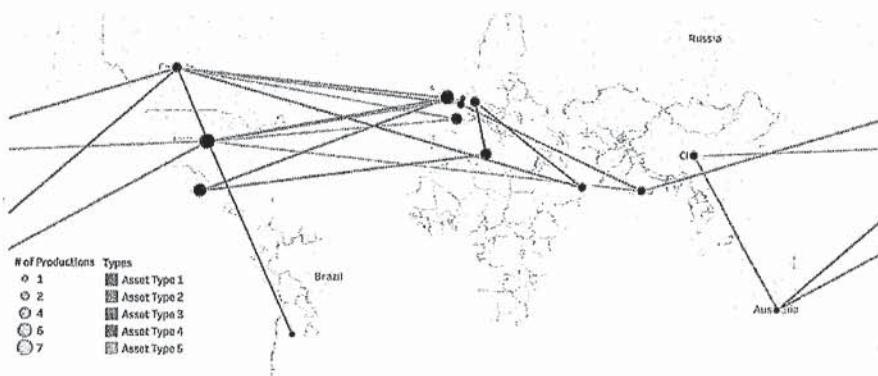


Giving users the ability to choose different filters (production team, geography, facility, etc.) and track trends over time builds an understanding of workflow patterns we might expect in the future, helping us plan more efficiently. Continuing our Sankey example, as we move toward Q4 it looks like the gaps in Blocked, In Progress, and Received even out and we've added more assets to the workflow. However, we might want to review the Q4 gap following the 'In Review' stage. This could lead teams to increase their staff, focus on technical improvements, or rebalance workload toward the end of the year.



As Netflix content production expands globally, geographic visuals are increasingly becoming an important part of analyzing studio data. By overlaying multiple dimensions across geographies, we can surface key insights on resource availability and dependencies, unlocking efficiencies across our production catalog.

A map like the one below might help teams better anticipate expected production delivery patterns for specific types of assets. Circles represent locations involved in production, and lines represent the movement of assets between locations.



Locations with larger circles and a high number of edges will experience a high influx of deliveries and may need extra planning to ensure all supporting resources are in place. For Assets with complex delivery patterns (e.g. Asset Type 4 above), we may have an opportunity to further optimize.

At the end of Post Production, if things go just right, we have a show or a movie ready for the screen. Quite importantly, at Netflix, this metaphorical screen is truly global. Millions of members, across 190 countries consume our content across over 20 languages. Thus

localizing content to make it ‘travel’ across the globe is an important part of Studio Production. We examine it next.

Ich liebe Netflix!

Localizing content originally created in, let’s say, German for the English-speaking US market is a complex creative process. Doing it well so that the content connects emotionally with viewers across the globe requires great effort and time.

Our localization teams craft an in-depth plan to create a seamless localized experience for our content. This includes developing a relationship with the content, assessing localization complexities, and providing specific creative guidance. For example, we try to cast voice actors in each language that sound like the original cast.

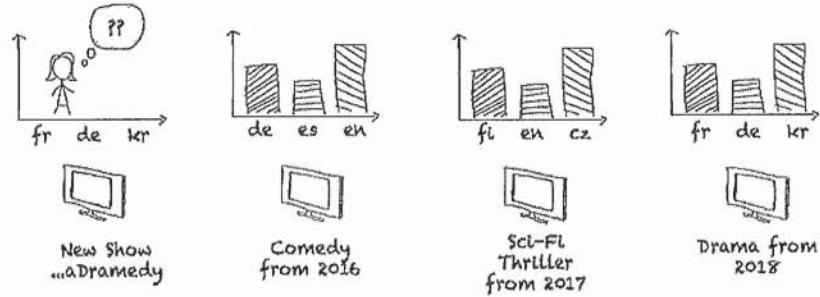


We strive to make all our content accessible to as many viewers as possible. Not surprisingly, time, talent availability, and technical constraints force us to sequence the available localization resources carefully. To make matters more complex, these decisions often need to be made many months before content is released on our service.

We can turn to data to support such decisions. Historical viewing trends inform us how our content is consumed across a range of languages and markets. If a piece of content is more popular in a language A than language B, we may sequence our efforts for A before B. For upcoming shows, this turns into the following data science problem: predict the per-language consumption for each show k months before it is released.

How do we solve this prediction problem? If we can define a notion of distance between shows, we can leverage historical data to make predictions about a new show’s per-language consumption based on

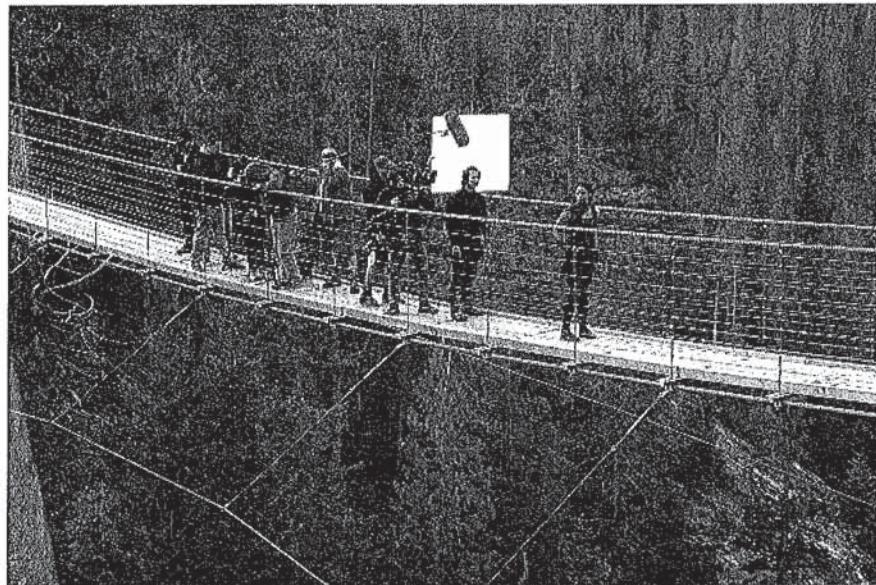
the per-language consumption of “similar” shows. Some features we might consider when constructing such a distance metric include genre, language (both the original language of the content as well as the localized language), and whether the localized content was consumed as dubbed audio or as subtitles.



Even a simple predictive model, as outlined here, can be helpful for guiding and scaling our content localization efforts as we expand our slate of localized content.

The typical next stop in a show’s journey towards launch (Figure 1) is Quality Control (QC). Managing QC workflows for multimodal (audio, video & text) creative products like scripted or unscripted shows, films, documentaries, etc. is a challenging task. Data Science plays an important role here by optimizing QC workflows using predictive modeling. You may read more about the specifics in our [previous blog post](#) on this subject.

Finally, we have a movie or TV show that’s ready to launch on Netflix! We’ve looked at only a few specific examples of problems where data science can help during the Studio Production stages, but the possibilities are limitless as Netflix ramps up original content creation.



Altered Carbon (Photo Credit: Katie Yu / Netflix)

Epilogue

It is not often that one gets to witness transformation of an entire industry. Opportunity to be an agent for that change is even rarer. Netflix has been that agent on a few different occasions over its short history. We believe we are at the cusp of another such transformation in the world of content creation. Working with Netflix Studio's business, technical and creative partners to transform a century-old industry with data science is challenging, but truly invigorating. If you are interested in being part of this refreshingly new endeavor with data, please contact [Ritwik Kumar](#) or check out the [Studio Production Science & Analytics](#) positions on the Netflix jobs site.

